A new human delayed-matching-to-place test in a virtual environment reverse-translated from the rodent watermaze paradigm: characterization of performance measures and sex differences

Matthew G. Buckley [a], Tobias Bast [b]

[a] Durham University, UK

[b] School of Psychology and Neuroscience@Nottingham, University of Nottingham, UK

Submitted to Hippocampus, 2017.

Please address correspondence to:

Tobias Bast, School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
tobias.bast@nottingham.ac.uk

Or

Matt Buckley, Depatment of Psychology, Durham University, Science Laboratories, Durham, DH1 3LE, UK
matthew.buckley@durham.ac.uk

Key words: spatial, learning and memory, hippocampus, water maze, translational

Abstract

Watermaze tests of place learning and memory in rodents, and corresponding reverse-translated human paradigms in real or virtual environments, are key tools to study hippocampal function. In common variants, the animal or human participant has to find a hidden goal that remains in the same place over many trials, allowing for incremental learning of the place with reference to distal cues surrounding the circular, featureless maze. Although the hippocampus is involved in incremental place learning, rodent studies have shown that the delayed-matching-to-place (DMP) watermaze test is a more sensitive assay of hippocampal function. On the DMP test, the goal location changes every 4 trials, requiring the rapid updating of place memory. Here, we developed a virtual DMP test reverse-translated from the rat watermaze DMP paradigm. In two replications, participants showed 1-trial place learning, evidenced by marked latency and path length savings between trials 1 and 2 to the same goal location, and by search preference for the vicinity of the goal when trial 2 was run as probe trial (during which the goal was removed). Performance was remarkably similar to rats' performance on the watermaze DMP test. In both replications, male participants showed greater savings and search preferences compared to female participants. Male participants also showed better mental rotation performance, although mental rotation scores did not consistently correlate with DMP performance measures, pointing to distinct neurocognitive mechanisms. The remarkable similarity between rodent and human DMP performance suggests similar underlying neuro-psychological mechanisms, including hippocampus dependence. The new virtual DMP test may, therefore, provide a sensitive tool to probe human hippocampal function.

1 INTRODUCTION

Translational behavioural assays using similar procedures in humans and animal models to measure distinct cognitive functions are important to reveal the causal neurobiological mechanisms underlying these functions and their impairments in neuropsychiatric disorders. One important approach involves the translation of human assays to animals, which led, for example, to the development of important rodent assays of clinically relevant cognitive functions (Brady & Floresco, 2015; Brown & Tait, 2016; Robbins, 2002), including the rodent touch screen battery to assess attention and memory (Hvoslef-Eide et al., 2015). A complementary reverse-translational approach involves adapting well-established rodent assays for testing in humans.

A prominent example of successful reverse-translation is the development of human assays of hippocampus-dependent memory function based on watermaze tests, which have long been a key tool to study hippocampus-dependent place learning and memory in rodents (Morris, 2008). The watermaze is a circular pool of opacified water, containing a submerged platform onto which the rat or mouse can escape. In the original and most common task variant, the platform remains in the same place over many trials and days of training, permitting the animals to incrementally learn the place of the hidden platform with reference to distal cues surrounding the watermaze (Morris, 1981). Place memory of where the platform is located in relation to the distal cues (i.e., allocentric place memory) is reflected by relatively short latencies and direct paths to the goal when the animals are placed into the pool from different start positions (which discourages use of egocentric strategies), and by persistent searching around the goal location when the platform has been removed for a probe trial. Performance on the incremental watermaze task in rats and mice is impaired by hippocampal lesions and by blockade of hippocampal plasticity mechanisms (Logue, Paylor, & Wehner, 1997; Morris,

Anderson, Lynch, & Baudry, 1986; Morris, Garrud, Rawlins, & O'Keefe, 1982; Morris, Schenk, Tweedie, & Jarrard, 1990; Morris, Steele, Bell, & Martin, 2013; Nakazawa, McHugh, Wilson, & Tonegawa, 2004; Sutherland, Whishaw, & Kolb, 1983; Tsien, Huerta, & Tonegawa, 1996). The incremental watermaze paradigm has been reverse-translated into human paradigms, in which participants are required to find a hidden goal in a real or virtual environment. Performance on these incremental watermaze task analogues is impaired in patients with hippocampal damage (Astur, Taylor, Mamelak, Philpott, & Sutherland, 2002; Barkas et al., 2012; Goodrich-Hunsaker, Livingstone, Skelton, & Hopkins, 2010; Nedelska et al., 2012), and has been used to probe impaired hippocampal function in various human conditions, including schizophrenia (Fajnerová et al., 2014; Folley, Astur, Jagannathan, Calhoun, & Pearlson, 2010; Hanlon et al., 2006; Rodriguez, 2010), clinical and non-clinical age-related decline (Daugherty, Bender, Yuan, & Raz, 2016; Daugherty et al., 2015; Driscoll, Hamilton, Yeo, Brooks, & Sutherland, 2005; Gazova, Vlcek, Laczó, Nedelska, Hyncicova, Mokrisova, Sheardova, & Hort, 2012; Hort, Laczo, Vyhnalek, Bojar, Bures, & Vlcek, 2007), as well as to characterise the development of hippocampal place memory in children (e.g. Balcomb, Newcombe, & Ferrara, 2011).

However, although the hippocampus is involved in incremental place learning, rats with complete hippocampal lesions can display intact place memory following extended incremental watermaze training (Bast, Wilson, Witter, & Morris, 2009; Morris et al., 1990; Whishaw & Jarrard, 1996), and partial hippocampal lesions, sparing less than half of the hippocampus, can leave performance on incremental watermaze paradigms relatively unaffected (de Hoz, Knox, & Morris, 2003; de Hoz, Moser, & Morris, 2005; Moser, Moser, Forrest, Andersen, & Morris, 1995). Moreover, rats can show intact incremental place learning in the watermaze even though hippocampal synaptic plasticity is blocked, if they have received pre-training (Bannerman, Good, Butcher, Ramsay, & Morris, 1995; Hoh, Beiko, Boon, Weiss,

& Cain, 1999; Inglis, Martin, & Morris, 2013; Otnæss, Brun, Moser, & Moser, 1999; Saucier & Cain, 1995). These results may reflect that incremental place learning, although normally facilitated by hippocampal mechanisms, can at least partly be supported by neo-cortical areas, whereas only rapid place learning may absolutely require hippocampal mechanisms (Bast, 2007; O'Reilly & Rudy, 2001).

Rapid place learning can be assessed using the delayed-matching-to-place (DMP) version of the watermaze test, which measures rodents' ability to learn within one trial the daily changing place of a hidden platform (Bast et al, 2009; Morris, Hagan, & Rawlins, 1986; Panakhova, Buresova, & Bures, 1984; Steele & Morris, 1999; Whishaw, 1985). The DMP task requires the continuous rapid updating of place memory, resembling the everyday task of remembering where we parked our car on a particular occasion. A common DMP watermaze protocol consists of daily 4-trial blocks to one location, with the hidden platform moving to a new location between days, i.e. successive 4-trial blocks (Bast et al., 2009; Steele & Morris, 1999). One-trial place learning is reflected by marked latency and path length savings, i.e. a steep reduction in these measures from trial 1 to 2, with little further improvements on subsequent trials, and – in more recent versions (Bast et al., 2009) – by a marked search preference for the vicinity of the correct location when trial 2 is run as probe, with the platform removed. Such one trial place learning performance, indexed by these measures, is highly sensitive to hippocampal dysfunction that may leave incremental watermaze performance relatively intact. More specifically, one trial place learning performance on the watermaze DMP test is severely impaired, and often virtually abolished, by complete and partial hippocampal lesions (Bast et al., 2009; de Hoz et al., 2005; Morris et al., 1990), as well as by disruption of hippocampal plasticity mechanisms (Inglis et al., 2013; Nakazawa et al., 2003; O'Carroll, Martin, Sandin, Frenguelli, & Morris, 2006; Pezze & Bast, 2012; Steele & Morris, 1999) and neuronal firing patterns (McGarrity, Mason, Fone, Pezze, & Bast, 2017). Therefore,

the DMP paradigm is a more sensitive assay of hippocampal function than incremental place learning paradigms.

An interesting recent paper (Fajnerova et al., 2014) reported a human virtual DMP assay, sharing some features with the rodent paradigm. During an 'acquisition' phase, participants had to learn a sequence of three different goal locations, completing 3 successive trials to each goal location, i.e. 9 trials overall. This resembles the rodent DMP paradigm, and performance measures showed a steep improvement between trial 1 and 2 to the same goal location, indicating 1-trial place learning similar to rodent studies. However, the performance measures (pointing error and path efficiency, i.e. ratio of direct to actual path to goal) were different from rodent studies and, in contrast to recent version of the rat DMP assay, trial 2 was never run as a probe. Consequently, 1-trial learning was not assessed in terms of search preference for the goal, which is the measure generally considered to be most reflective of a hippocampus-dependent allocentric representation of the goal location in relation to multiple distal spatial cues (Burešová, Krekule, Zahalka, & Bureš, 1985; Jacobs & Schenk, 2003; Morris, 1981; Schenk & Morris, 1985) and which, in line with this, has been demonstrated to be most sensitive to hippocampal manipulations on the rat DMP test (Bast et al., 2009; Pezze & Bast, 2012; McGarrity et al., 2017) (see **4.1** for a more detailed discussion of differences between search preference and other DMP performance measures). Moreover, the only spatial cues were three landmarks within the circular testing arena, whereas in the rodent paradigm the cues for goal localization are outside the pool, and the size of the goal locations made up 10% of the whole testing arena, making goal localization substantially easier than in the rodent task, where the goal locations occupies <1 % of the pool area. Overall, whilst the study by Fajnerova et al. (2014) supports the feasibility of a human DMP test, the differences outlined above limit direct comparability to the rodent paradigm.

Here, we report the development of a new human virtual DMP test, closely adapted from our rat watermaze DMP paradigm (Bast et al., 2009; da Silva, Bast, & Morris, 2014; McGarrity et al., 2016; Pezze & Bast, 2012). Within a virtual environment presented on a computer screen, participants were required to find a hidden goal that remained in the same location for four consecutive trials, after which the goal moved to a new location (**Fig. 1**). Based on the rat studies, we expected participants to show 1-trial learning, reflected by marked latency and path length savings from trial 1 to 2, as well as a pronounced search preference for the vicinity of the goal when trial 2 was run as probe. To allow for a direct comparison of human and rat data, we retrospectively analysed data from a large sample of rats from our previous watermaze DMP experiments, using the same analyses as for the human data.

As part of the first characterisation of performance in our novel DMP task, we also assessed sex differences. Many previous studies have shown that male rats or male human participants show better place learning and memory than female rats or female human participants on incremental watermaze paradigms (e.g., Keeley, Tyndall, Scott, & Saucier, 2013; Saucier, Shultz, Keller, Cook, & Binsted, 2008; see Jonasson et al., 2005, for a review) or on corresponding human paradigms (Astur, Ortiz, & Sutherland, 1998; Astur, Purton, Zaniewski, Cimadevilla, & Markus, 2016; Astur, Tropp, Sava, Constable, & Markus, 2004; Driscoll et al., 2003; Leon, Cimadevilla, & Tascon, 2014; Woolley et al., 2010; Padilla, Creem-Regehr, Stefanucci, & Cashdan, 2017). However, there is some heterogeneity, and not all studies report significant sex differences in rats (Bucci, Chiba, & Gallagher, 1995; Faraji, Metz, & Sutherland, 2010) or humans (Daugherty et al., 2016; Hamilton et al., 2009; Moffat & Resnick, 2002; see also Driscoll et al., 2005, in which male participants did not perform better than female participants in terms of search preference, although they did perform better in terms of latencies). Studies in mice have produced a particularly heterogenous picture, with findings of a male advantage, a female advantage, and of no significant sex difference on incremental

watermaze paradigms; aggregate analyses across studies and involving large samples revealed only negligible effect sizes (Cohen's $d < 0.2$) (see Jonasson, 2005; Fritz, Amrein, & Wolfer, 2017; and references therein). Moreover, some authors suggested that some assays of hippocampus-dependent place learning and memory may not show sex differences in rats or human participants (Astur et al., 2004; Faraji et al., 2010). Fajnerova et al. (2014) reported that male, as compared to female, participants tended ($p=0.06$) to show overall increased path efficiency during the acquisition phase of their task, which resembles our new DMP test, although the critical interaction between sex and trial to the same goal location, which would support sex differences in 1-trial place learning, was not reported. In the present study, we might expect greater latency and path length savings between trial 1 and 2, as well as higher search preference during probe trials in male as compared to female participants on the DMP test, based on the weight of previous evidence, although this is not a forgone conclusion. A factor that is likely to have contributed to the heterogeneity in previous studies of sex differences is limited statistical power, which increases the probability of both false negatives and false positives (Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson, & Munafo et al., 2013). In the present experiments, we recruited large enough samples for two appropriately powered replications to ensure that any findings, including sex differences, were reliable (see **2.1.** for details).

Finally, for comparison, participants also completed the Vandenberg and Kuse (1978) mental rotation test (MRT). This classic assay of human visuo-spatial ability has quite reliably revealed that, on average, male participants perform better than female participants (Astur et al., 2004; Peters et al., 1995; Voyer, Voyer, & Bryden, 1995), resembling the picture emerging for place learning on watermaze analogues. However, MRT performance has mainly been associated with cortical visuo-motor regions (Kosslyn, Digirolamo, Thompson, & Alpert, 1998; Kosslyn, Ganis, & Thompson, 2001), whereas place learning on watermaze analogues

has mainly been linked to the hippocampus. Therefore, whilst we expected mental rotation and DMP measures to reveal similar male performance advantages, they may show only limited statistical correlation, supporting that both assays measure distinct neuro-cognitive processes (but see Astur et al., 2004, and Driscoll et al., 2005).

2 METHODS AND MATERIALS

2.1 Participants and overall study design

Based on previous studies showing sex differences in place learning using an incremental multi-trial learning protocol in a virtual watermaze test (e.g. Astur et al., 2004), and in mental rotation (Voyer et al., 1995), we expected medium effect sizes (Cohen's d = 0.5, η_p^2 = .06) for the sex difference in search preference and in mental rotation scores. G*Power 3.1.9.2 (Faul et al., 2007) was used to determine that a sample size of 128 participants (64 male, 64 female) was required to detect a medium effect size (d = 0.5) when using a 2-tailed t-test, with an alpha level of 0.05 and power of 0.8.

Two replications of this study were run, one year apart, with two different groups of experimenters, who were undergraduate Psychology students at the University of Nottingham, collecting the data for their final-year theses (in the 2014/15 and 2015/16 academic years). Each experimenter tested a similar number of male and female participants, to minimize the risk that experimenter characteristics, such as sex (Chapman, Benedict, & Schioth, 2017), confounded the study of sex differences. For Replication 1, four experimenters (2 female, 2 male) recruited a total of 123 participants (60 female), who were aged between 18 and 33 years (mean = 20.50 years, SD = 2.02 years). For Replication 2, five experimenters (2 female, 3 male) recruited 133 participants (68 female), who were aged between 18 and 30 years (mean =

19.88, SD = 1.81). Participants were mainly students of the University of Nottingham and recruited directly by the experimenters from personal acquaintances or through the School of Psychology Online Research Participation Scheme, which offers Psychology undergraduate students course credit in return for participation.

After giving informed consent and completing a brief questionnaire about participant demographics (typically 5-10 min), participants completed the new virtual DMP test (typically 15-30 min), followed by the Vandenberg and Kuse Mental Rotation Test (typically about 5 min of instruction based on example sets, plus 10 min for the test).

2.2 Apparatus and materials

Virtual DMP test. The virtual environment was presented on an Apple Mackintosh model A1224 (EMC2133) with a screen of 27.40 x 43.40 cm and consisted of a circular lawn area enclosed by a wooden fence, around which 8 different distal landmark cues were arranged (**Fig. 1 A, C**). All virtual environments were constructed and displayed using Mazesuite (v2.1) software (Ayaz, Allen, Platek, & Onaral, 2008; www.mazesuite.com). The virtual environments were presented to the participants from a first-person perspective with a field of view of 45 degrees. The size of the virtual environment in relation to the human participant moving in the virtual environment was chosen to be comparable to the size of the watermaze in relation to the rat swimming in the watermaze in the studies by Bast and colleagues (Bast et al., 2009; da Silva, Bast, & Morris, 2014; McGarrity et al., 2016; Pezze & Bast, 2012). More specifically, in the watermaze DMP test, rats swim at about 20 cm/s in a 2 m diameter pool, i.e. can travel the whole diameter within about 10 s, whilst searching for a submerged circular platform of 12 cm diameter, i.e. a surface area that is about 0.3% of the size of the pool surface. Therefore, in our virtual watermaze built in Mazesuite, the diameter of the wall that enclosed

the circular environment was chosen such that participants could travel the whole diameter within 10 s by pressing the forward arrow key on the computer, resulting in a diameter of 26.77 MazeUnits (Mu, the measure of virtual distance in MazeSuite). The wall was 2 Mu tall. A wooden fence texture was applied to the wall of the environment, a grass texture was applied to the floor, and the sky was rendered a uniform black expanse. Goal locations within the environment were square-shaped with side-lengths of 1.8 Mu, corresponding to a surface area of about 0.14 % of the size of the circular virtual environment, and invisible to participants. Following the rodent watermaze DMP studies, the goal locations were chosen from a set of eight locations that were evenly distributed across the pool, with the centre of the goals aligned with an inner (diameter 12.17 Mu) or an outer (diameter 18.44 Mu) ring that was concentric with the arena walls (**Fig. 1A**). Eight unique landmark cues were placed at varying distances away from the circular wall. The first landmark was placed 22.5° clockwise of the notional north (N) of the environment, and each successive landmark was placed a further 45° clockwise (**Fig. 1A, C)**. The objects were the following distances from the centre of the arena: hot-air balloon 15.00 Mu, space shuttle 20.56 Mu, tree 14.74 Mu, Hubble telescope 26.44 Mu, star 20.09 Mu, castle tower 15.11 Mu, planet 20.90 Mu, wind-turbine 15.85 Mu. All models were sourced from turbosquid.com, except for the Hubble telescope and space shuttle, which were sourced from nasa.gov.

Mental rotation test. We used the standard version MRT-A of the Vandenberg and Kuse Mental Rotation Test, with stimulus figures redrawn from the original Vandenberg & Kuse (1978) set and containing 24 problem sets (Peters et al., 1995).

2.3 Procedure

All testing was completed in the same quiet testing room in the School of Psychology, with the experimenter present.

Virtual DMP test. Participants received the following instructions on paper:

You will complete a spatial learning computer task in a virtual environment consisting of a circular lawn surrounded by a fence. Your task is to find William the Worm, who is hidden under the grass (you can't see him), by moving across the virtual lawn.

William will remain in the same location for 4 trials in a row, after which you will be notified on the computer screen that William has moved to a new location. This will happen 6 times, and in total you will complete 24 trials. If you cannot find William after 2 minutes then a white flag will appear showing his location, which you must move towards. Importantly, you can see several objects located around the outside of the fence. These objects can help you learn exactly where William is hidden, helping you to find him more quickly. It is normal that you might find this task difficult.

Each trial will begin facing the fence, at one of four start positions (North, East, South or West). You will need to use the arrow keys on the keyboard to turn around and move across the virtual lawn so that you can find William.

When you find William you will be notified by a message on the computer screen, and a picture of William will appear.

The William-the-Worm cover story used here has been used previously, and is readily understood by both young children and adults (Buckley, Haselgrove, & Smith, 2015). When completing the virtual DMP test, participants sat approximately 50 cm from the screen, and navigated through the virtual environment using the cursor keys. Presses on the "up" and "down" cursor keys permitted the participant to move forwards and backwards within the

arena, respectively, while presses on the "left" and "right" cursor keys permitted the participant to rotate counter-clockwise and clockwise within the environment, respectively.

Procedures for DMP testing were adapted from the rodent paradigm (Bast et al., 2009; da Silva et al., 2014; McGarrity et al., 2017; Pezze and Bast, 2012; Steele & Morris, 1999). During the navigation task, the hidden goal remained in the same location for only four consecutive trials, after which the goal was moved to a new location for another block of four trials. This follows the standard procedure of the rat watermaze DMP paradigm, although the rats typically only complete one four trial block per day, with goal locations changing between days. In order to prevent the use of egocentric strategies to navigate to the goal, participants began each of the four trials within a block at one of four different start locations spaced evenly along the fence perimeter (N, E, S, W – see **Fig. 1A**). The order of these start locations was arbitrary, and the same for each participant. Throughout the experiment, all participants received the same sequence of 6 different goal locations. The sequence of goal locations and starting positions were as follows: Location 1 (WENS), Location 2 (SNEW), Location 3 (EWSN), Location 4 (NSEW), Location 5 (ENSW), Location 6 (NWSE) (for locations, see **Fig. 1A**). The task was self-paced, with the participant determining the start of each trial by pressing the enter button (typically, within about 5 s of the last trial), whereas in the standard rat DMP paradigm the animals are removed from the testing environment between trials and replaced in the pool by the experimenter, with a minimum inter-trial interval of 15-30 s. As indicated in the participant instructions above, at the start of each trial, participants faced the fence at the given start location and had to turn away from the fence using the arrow keys (similar to the situation in watermaze tests, where the animals are placed into the pool facing the pool wall).

For 22 out of the 24 trials, the hidden goal was present. On navigating to the goal on these acquisition trials, participants could no longer move and received a congratulatory message (*You found William! Press enter*). On pressing enter, a picture of a cartoon worm was presented for 1s, after which the next trial began automatically. If participants had not found the goal within 120 s on an acquisition trial, a white flag appeared at the goal location within the environment. As noted before, the goal was positioned in the same location for 4 consecutive trials, after which it moved to a new location. Prior to the first trial at a new goal location, participants were informed that the goal was in a new position by displaying a message (*William has moved to a new location*) on screen for 3 seconds.

On two occasions, trial 2 at a given hidden goal location was conducted as probe trials. During such probe trials, unbeknownst to the participants, the hidden goal was removed from the environment, to enable the measurement of search preference for the correct zone containing the goal location (see below, Performance measures). Participants were allowed to search for 60 s, after which the trial terminated automatically. Following Buckley et al. (2016), when the probe trial terminated, a message (*Keep looking for William!*) was displayed on screen for 1 second, and then the next acquisition trial (i.e., trial 3) began automatically. The two probe trials were always conducted on the second trial to goal location 4 and 6.

Virtual DMP performance measures. Following studies using the rat watermaze DMP (Bast et al. 2009; da Silva et al., 2014; McGarrity et al., 2017; Pezze & Bast, 2012), search preference for a small zone surrounding the goal location during the 60-s probe trials was the main measure of 1-trial place memory. To quantify search preference, eight square (5.4x5.4 Mu) zones were defined on the inner and outer ring of the circular arena, so that one zone (the correct zone) was concentric with the goal location, and all eight zones were non-overlapping, evenly spaced and symmetrically arranged over the remaining goal locations. The

time spent in each of these eight zones during the 60-s probe trial was determined automatically using the MazeSuite software. From these measures, the percentage of time spent in the correct zone was calculated as: (time in correct zone [s] / time in all eight zones [s]) x 100. By chance, this value should be 100 % / 8 = 12.5 %, whereas higher values indicate a search preference for the correct zone.

In addition, latencies and path lengths to enter the goal location (or, on probes, the location where the goal would have been) were recorded for all trials, with steep reduction from trial 1 to 2 indicating 1-trial place memory. We expected latencies and path lengths to show similar patterns of changes across trials. In contrast to latencies, however, path lengths measure the efficiency in reaching the platform independent of speed (e.g., Bast et al., 2009) and are, therefore, less susceptible to potential differences in participants' sensorimotor proficiency in using the arrow keys. On probe trials, if a participant did not enter the location where the platform should have been, the maximum trial duration (60 s) was recorded for latency, and the total distance traversed during this time was recorded as path length.

Mental rotation test. After completing the virtual DMP test, participants completed the Vandenberg and Kuse Mental Rotation Test, using the standard version MRT-A with stimulus figures redrawn from the original Vandenberg and Kuse (1978) set (Peters et al., 1995). This task comprises 24 problem sets with a target figure on the left of the page, and four similar figures on the right. Participants were required to mark the two figures on the right that were rotated versions of the target figure, and leave blank the remaining two figures that did not match the target figure. Following Peters et al. (1995), participants first received specific instructions on the principle of the test and were asked to complete four example problems followed by feedback; then, they were allowed 3 min to complete as many of the first 12 problems as possible, given a 4-min break, and then allowed a further 3 min to complete as

many of the second 12 problems as possible. The test was scored by counting the number of correct responses made by a participant. A response to a given question was considered correct only if both the rotated figures were identified, yielding a maximum score of 24.

2.4 Statistical analysis

For all statistical analyses, an alpha level of $p<0.05$ was accepted as significant.

Virtual DMP performance measures. Latencies and path lengths for trials 1 to 4 were averaged across all six locations. Both latency and path length data were then subjected to analysis of variance (ANOVA), with factors of sex and trial. In addition, within each sex, we used paired-samples t-tests to test for a significant reduction of latencies and path lengths from trial 1 to 2, reflecting significant one-trial place learning. To assess if these trial 1 to 2 savings were greater in male than female participants, we subtracted trial 2 performance from trial 1 performance, and the resulting latency and path length savings were compared using a between-subjects t-test.

Search preference data from the two probe trials, expressed as percentage of time spent in the correct zone, were subjected to ANOVA with sex as between-subjects and probe trial as repeated measures factor. In addition, to demonstrate significant one-trial place memory, we compared the percentage of time spent in the correct zone to chance (12.5%) using one-sample t-tests.

To address if the key performance measures on virtual DMP test reflect related or partially dissociable neuro-psychological mechanisms, we examined the linear associations between search preference, latency and path lengths savings by calculating Pearson's product-moment coefficients for pairs of these measures.

Mental rotation scores. Means were compared between sexes, using independent t-tests.

Correlation between performance measures on the virtual DMP task and on the mental rotation test. To address if key performance measures on the virtual DMP test and the mental rotation scores reflect related or partially dissociable neuro-psychological mechanisms, we examined the linear associations between search preference, latency and path lengths savings, and the mental rotation score by calculating Pearson's product-moment coefficients for pairs of these measures.

Impact of computer gaming. In the participant demographic questionnaires, we asked all participants if they 'regularly play computer games'. Across the two replications, 83 male and 10 female participants indicated that they regularly played computer games, i.e. there was a substantial sex imbalance (also compare Terlecki & Newcombe, 2005). Therefore, we wished to address if regular computer gaming may have contributed to sex differences in performance measures on the virtual DMP and mental rotation test. Unfortunately, it was not possible to analyse data using sex and gaming experience as separate factors in an ANOVA due to the very low number of female gamers. Instead, we conducted the following analyses on combined samples from both replications: first, in the combined male sample, we compared search preference and latency and path length savings on the DMP test, as well as the mental rotation score between gamers (83 male) and non-gamers (45 male), using between-subjects t-tests; second, we compared sex differences in these measures in the combined sample of non-gamers, i.e. following exclusion of all gamers, which left 118 female and 45 male non-gamers. It should be noted that these additional analyses were underpowered as compared to the primary analyses of sex differences due to the low numbers of female gamers and male non-gamers in both replications, and, therefore, the outcomes need to be considered with caution.

2.5 Comparison to data from the rat DMP test in the watermaze

For comparison, we re-analysed data collected from a large sample of rats that underwent pre-training on the watermaze DMP test before they underwent surgery for a lesion study (Bast et al., 2009). The data are from 100 male young adult Lister hooded rats that completed watermaze DMP training consisting of 4 trials each to 8 different locations. In contrast to human testing (where all testing is completed within one session), rats were tested across 8 days, with 4 trials to one location per day. Moreover, while the human DMP test is self-paced, rats were tested with a trial 1-to-trial 2 delay of 15-30 s or 20 min (each retention delay was used equally often, i.e. on 4 days), with all other inter-trial intervals 15-30 s. On days 4 and 8, trial 2 was run as probe, with the target platform not accessible for a 60-s period (the trial 1-to-trial 2 delay was 15-30 s for one probe trial and 20 min for the other one). Average latencies across trials 1-4, average trial 1 and 2 latency savings and average search preference during probes, as well as the correlations between these measures, were analysed as described for the human data above.

3 RESULTS

3.1 Virtual DMP test: robust one-trial place learning in both sexes, with male participants performing better than female participants

Visual inspection of paths during trial 1 to 4. Paths during trial 1 were typically characterised by systematic searching, whereas participants moved more directly towards the goal location during the subsequent trials (2-4), reflecting rapid place learning (**Fig. 2**). During trial 1, participants most commonly searched in a circular pattern, not so unlike the pattern

typically shown by rats (e.g., see Fig. 6 in Steele & Morris, 1999), although other patterns were also observed in some participants, including zig-zag patterns.

Latencies. In both replications, both male and female participants showed marked one-trial place learning, reflected by steep reductions in the latency to find the goal from trial 1 to 2, with little further latency reductions on trials 3 and 4 (**Fig. 3**). Male participants also tended to find the goal quicker than female participants across all trials and tended to show greater savings from trial 1 to 2, suggesting better 1-trial place learning.

For Replication 1, ANOVA revealed a significant main effect of sex ($F_{(1, 121)} = 45.56$, $MSE = 514.70$, $p < .001$, $\eta_p^2 = .27$), trial ($F_{(3, 363)} = 120.04$, $MSE = 293.47$, $p < .001$, $\eta_p^2 = .50$), and a significant interaction between sex and trial ($F_{(3, 363)} = 3.21$, $MSE = 293.47$, $p = .023$, $\eta_p^2 = .03$). Post-hoc t-tests between sexes revealed that male participants were significantly quicker than female participants to find the goal on trial 2 ($t_{(121)} = 4.96$, $p < .001$, $d = .90$), trial 3 ($t_{(121)} = 3.15$, $p = .002$, $d = .57$), and trial 4 ($t_{(121)} = 7.41$, $p < .001$, $d = 1.34$) and also strongly tended to find the goal more quickly on trial 1 ($t_{(121)} = 1.97$, $p = .051$, $d = .36$). Both male ($t_{(62)} = 14.59$, $p < .001$, $d = 2.27$), and female participants ($t_{(59)} = 6.08$, $p < .001$, $d = 1.16$) were significantly quicker to find the goal on trial 2 compared to trial 1. Savings, i.e. the difference between trial 1 and 2 latencies, were greater in male than female participants ($t_{(121)} = 2.15$, $p = .033$, $d = .39$).

For Replication 2, there was a significant main effect of sex ($F_{(1, 131)} = 45.54$, $MSE = 22623.25$, $p < .001$, $\eta_p^2 = .26$), trial ($F_{(3, 393)} = 289.18$, $MSE = 59970.66$, $p < .001$, $\eta_p^2 = .69$), and a significant interaction between sex and trial ($F_{(3, 393)} = 4.08$, $MSE = 849.75$, $p = .007$, $\eta_p^2 = .03$). Post-hoc t-tests between sexes revealed that male participants were significantly quicker than female participants to find the goal on trial 2 ($t_{(131)} = 7.23$, $p < .001$, $d = 1.25$), trial 3 ($t_{(131)} = 4.87$, $p < .001$, $d = .84$), trial 4 ($t_{(131)} = 4.42$, $p < .001$, $d = .78$), and also on trial 1 ($t_{(131)} = 7.23$, $p = .04$, $d = .35$). Within sexes, both male ($t_{(64)} = 17.50$, $p < .001$, $d = 3.04$) and female ($t_{(67)} =$

13.14, $p<.001$, $d = 1.92$) participants were significantly quicker to find the goal on trial 2 compared to trial 1. Latency savings between trial 1 and 2 were greater in male than female participants ($t_{(131)} = 3.29$, $p = .001$, $d = .57$).

Path lengths. In keeping with latency data, path lengths showed a steep decrease from trial 1 to 2 for both male and female participants, reflecting marked one-trial place learning. Male participants also showed greater savings from trial 1 to 2 across both replications, relative to female participants (**Fig 3**).

For Replication 1, ANOVA of latencies revealed a significant main effect of sex ($F_{(1, 121)} = 9.68$, $MSE = 1348.54$, $p= .002$, $\eta_p^2 = .07$), trial ($F_{(3, 363)} = 173.73$, $MSE = 1115.27$, $p< .001$, $\eta_p^2 = .59$), and a significant interaction between sex and trial ($F_{(3, 363)} = 5.15$, $MSE = 1115.27$, $p=.002$, $\eta_p^2 = .04$). Post-hoc t-tests between sexes revealed that male participants travelled significantly shorter distances than female participants to find the goal on trial 2 ($t_{(121)} = 2.64$, $p=.009$, $d = .48$) and trial 4 ($t_{(121)} = 5.35$, $p<.001$, $d = .97$), but not on trial 1 ($t_{(121)} = .89$, $p=.38$, $d = .16$) or trial 3 ($t_{(121)} = 1.17$, $p=.243$, $d = .21$). Within sexes, both male ($t_{(62)} = 14.73$, $p<.001$, $d = 2.50$) and female ($t_{(59)} = 7.61$, $p<.001$, $d = 1.58$) participants travelled significantly shorter distances to find the goal on trial 2 compared to trial 1. In keeping with latency data, the path length savings from trial 1 to 2 made by male participants were significantly greater than those made by female participants ($t_{(121)} = 2.18$, $p = .031$ $d = .39$).

For Replication 2, there was a significant main effect of sex ($F_{(1, 131)} = 5.89$, $MSE = 1935.76$, $p= .043$, $\eta_p^2 = .04$), trial ($F_{(3, 393)} = 298.24$, $MSE = 956.01$, $p< .001$, $\eta_p^2 = .70$), and a significant interaction between sex and trial ($F_{(3, 393)} = 6.20$, $MSE = 956.01$, $p< .001$, $\eta_p^2 = .05$). Post-hoc t-tests between sexes revealed that male participants travelled significantly shorter distances to find the goal on trial 2 ($t_{(131)} = 4.87$, $p<.001$, $d = .85$) and trial 4 ($t_{(131)} = 2.39$, $p<.018$, $d = .42$) compared to female participants, but this was not the case on trial 1 ($t_{(131)} =$

1.39, p=1.67, d = .24) or trial 3 ($t_{(131)}$ = 1.52, p=.13, d = .26). Within sexes, both male ($t_{(64)}$ = 18.39, p<.001, d = 3.29) and female ($t_{(67)}$ = 15.17, p<.001, d = 2.19) participants were significantly quicker to find the goal on trial 2 compared to trial 1. Again, path length savings were greater in male participants compared to female participants ($t_{(131)}$ = 4.24, p <.001, d = .73).

Search preference. In both replications, both male and female participants showed strongly above-chance search preference for the correct zone during both probes, supporting one-trial place learning, and male participants showed higher search preference than female participants (**Fig. 3**).

For Replication 1, ANOVA of mean proportions of time spent in the correct zone during the two probes revealed a significant main effect of sex ($F_{(1, 121)}$ = 6.69, MSE = 879.71, p= .01, η_p^2 = .05), but the main effect of probe and the interaction between sex and probe were not significant (both $F_{(1, 121)}$ < 1). Although male participants showed higher search preference than female participants, both female and male participants spent more time in the correct zone than would be expected by chance (12.5%). To compare performance on probe trials to chance, the average proportion of time spent in the correct zone on each probe was compared to chance for each sex. On probe 1, both male ($t_{(62)}$ = 12.91, p<.001, d = 1.63) and female ($t_{(59)}$ = 7.60, p<.001, d = .98) participants spent more time searching in the correct zone than would be expected by chance, and the same was true for male ($t_{(62)}$ = 11.02, p<.001, d = 1.39) and female ($t_{(59)}$ = 7.99, p<.001, d = 1.03) participants during probe 2.

For Replication 2, there was again only a significant main effect of sex ($F_{(1, 131)}$ = 6.26, MSE = 917.43, p= .014, η_p^2 = .05). The main effect of probe, and the interaction between sex and probe were not significant (both $F_{(1, 131)}$ < 1). Again, both male and female participants seemed to spend more time in the correct zone than would be expected by chance. One-sample

t-tests conducted on individual average proportions of time traversed in the correct zone in the first probe revealed that male ($t_{(64)} = 13.32$, $p<.001$, $d = 1.65$) and female ($t_{(68)} = 9.69$, $p<.001$, $d = 1.17$) participants spent more time searching in the correct zone than would be expected by chance, and the same was true for male ($t_{(64)} = 13.00$, $p<.001$, $d = 1.61$) and female ($t_{(68)} = 9.35$, $p<.001$, $d = 1.13$) participants during probe 2.

To assess if participants also still expressed a search preference for the previous goal location on probe trials, we calculated the proportion of time that participants spent in the zone surrounding the previous location. For Replication 1, the proportion of time participants spent in the zone surrounding the previous location (mean$\pm$SEM) was 0.02$\pm$0.004 % and 0.03$\pm$0.007 % for probe 1 and 2, respectively. In Replication 2, these proportions for probe 1 and 2 were 2.4$\pm$0.59 % and 2.3$\pm$0.50 %, respectively. In all probe trials, therefore, participants spent significantly less time in the zone surrounding the previous location than would be expected by chance (all $t >17.20$, $p <0.001$, $d >1.50$).

Correlations between performance measures on the virtual DMP task. Not surprisingly, latency and path length savings showed a very strong positive correlation in Replication 1 ($r_{(121)} = .92$, $p<.001$) and Replication 2 ($r_{(131)} = .83$, $p<.001$) (data not shown). Moreover, if participants showed higher latency and path length savings from trial 1 to 2, they also tended to show higher search preference for the correct zone on probe trials. However, although highly significant, the correlations of latency and path length savings with search preference were only partial. Pearson product-moment coefficients revealed significant correlations between search preference (averaged across the two probes) and latency savings for both Replication 1 ($r_{(121)} = .41$, $p<.001$) and Replication 2 ($r_{(131)} = .25$, $p=.003$) (**Fig. 4**). Similarly, there were significant correlations between search preference (averaged across the two probes) and path length savings in Replication 1 ($r_{(121)} = .38$, $p<.001$) and Replication 2

($r_{(131)}$ = .37, p<.001) (data not shown). Based on these correlations, latency and path length savings can account for <17% of the variance in search preference.

Comparison to data from the rat DMP test in the watermaze. The data on the human virtual DMP test are remarkably similar to data collected from rats performing the watermaze DMP test, with similarly steep latency reductions from trial 1 to 2 and marked search preference for the correct zone during probes (**Fig. 5**). A one-way ANOVA of the rat latencies across trials 1 to 4, with a within-subjects factor of trial, revealed a significant main effect of trial ($F_{(3, 297)}$ = 469.82, *MSE* = 111.53, p< .001, η_p^2 = .83). Rats showed a very pronounced reduction in latencies between trial 1 and 2 ($t_{(99)}$ = 20.29, p = <.001, d = 2.78), as well as a marked search preference for the correct zone (as compared to chance, 12.5 %) during probe trials ($t_{(99)}$ = 15.74, p = <.001, d = 2.01). Finally, there was a partial correlation of similar strength as in humans between latency savings and performance during probe trials ($r_{(98)}$ = .51, p< .001).

We also calculated the proportion of time that rats spent in the zone surrounding the previous location to assess if memory for the previous goal location interfered with search preference. In contrast to human participants, who spent only negligible amounts of time in the zone surrounding the previous location, the proportion of time (mean$\pm$SEM) that rats spent in the zone surrounding the previous location during probes was 14.40$\pm$0.78 %, which was significantly higher than expected based on chance ($t_{(99)}$ = 2.46, p = 0.016, d = .25).

3.2 Mental Rotation: male participants perform better than female participants

Male participants scored higher on the mental rotation test than female participants in both Replication 1 ($t_{(121)}$ = 7.23, p<.001, d = 1.30) and Replication 2 ($t_{(131)}$ = 2.77, p=.006, d = .48) (**Fig 6**).

3.3 No consistent correlation between performance measures on the virtual DMP task and on the mental rotation test

Virtual DMP performance did not vary consistently with mental rotation scores (**Fig. 7**). Pearson product-moment coefficients revealed no significant correlations between search preferences (averaged across the two probes) and mental rotation scores in Replication 1 ($r_{(121)}$ = .13, p= .15) or Replication 2 ($r_{(131)}$ = .14, p= .10). Moreover, in Replication 1, average latency savings ($r_{(121)}$ = .09, p= .33) and path length savings ($r_{(121)}$ = .12, p= .19) did not correlate with mental rotation scores, although, in replication 2, there were weak, but significant, correlations of average latency savings ($r_{(131)}$ = .21, p= .015) and path length savings ($r_{(131)}$ = .25, p= .004) with mental rotation scores.

3.4 Impact of computer gaming on performance measures on the virtual DMP task and on the mental rotation test

Our participant demographics indicated that male participants were more likely to regularly play computer games than female participants. Across the two experiments 83 male and 10 female participants indicated that they regularly played computer games, whilst 118 female and 45 male participants indicated that they did not regularly play computer games. We therefore wished to assess the impact that gaming may have on our data. Unfortunately, though, it was not possible to analyse data using sex and gaming experience as separate factors in an ANOVA due to the low number of female gamers. Consequently, in order to estimate how strongly gaming impacts mental rotation and virtual DMP performance measures, we, first, combined the male samples from both experiments to examine the impact of gaming

experience, independent from any sex differences. Between-subjects t-tests on this combined sample revealed that gamers had significantly greater latency savings ($t_{(126)} = 2.70$, $p = .008$, $d = .48$), path length savings ($t_{(126)} = 2.39$, $p = .018$, $d = .43$), and higher mental rotation scores ($t_{(126)} = 2.01$, $p = .047$, $d = .38$) than non-gamers. Interestingly, however, there were no significant differences between gamers and non-gamers in the proportion of time spent within the correct zone during probe trials ($t_{(126)} = 1.38$, $p = .17$, $d = .26$). Overall, although there was a significant effect of gaming on virtual DMP and mental rotation performance measures, this effect seems to be less pronounced than the sex difference in these measures, so can – if at all – only partially account for these sex differences. Second, we assessed if there were sex differences in all non-gamers in our sample, across both replications. Between-subjects t-tests on this combined sample revealed no significant differences between male and female non-gamers in latency savings ($t_{(161)} = .81$, $p = .42$, $d = 0.14$), or path length savings ($t_{(161)} = 1.33$, $p = .19$, $d = 23$). However, male non-gamers scored higher on the mental rotation test compared to female non-gamers ($t_{(161)} = 3.48$, $p = .001$, $d = .62$), and the higher proportion of time spent within the correct zone during probe trials by male compared to female participants approached significance ($t_{(161)} = 1.90$, $p = .059$, $d = .34$).

These additional analyses, although underpowered as compared to the primary analysis of sex differences, support that male participants displayed higher mental rotation scores and DMP search preference than female participants, independent of gaming experience. However, they also suggest that gaming experience, independent of sex differences, may be associated with better DMP latency and path length savings, as well as better mental rotation scores. Taken together, these additional analyses suggest that more gaming experience in male participants may, if at all, only partially account for their better virtual DMP and mental rotation performance relative to female participants. It must also be noted that any conclusion concerning the causal role of gaming experience in changing performance measures in our

sample is limited by our simple self-report measure of gaming and our lack of assessment of other variables.

4 DISCUSSION

On our new virtual DMP test, participants were required to navigate to a hidden goal that was moved to a new location every four trials and whose location was defined by its relation to distal cues surrounding the circular environment. In two replications, participants showed clear 1-trial place learning, evidenced by marked latency and path length savings between trials 1 and 2, and by a search preference for the vicinity of the goal when trial 2 was run as probe. Performance was remarkably similar to rats' performance on the watermaze DMP test. In both replications, male participants showed greater savings and search preferences compared to female participants. Male participants also showed greater mental rotation scores, but mental rotation scores did not consistently correlate with DMP performance measures.

4.1 The new virtual DMP test

A remarkable feature of DMP performance in both rodents and humans is the evidence for repeated one-trial learning of new places within a familiar environment. As has been pointed out by others (Steele & Morris, 1999; Whishaw, 1985), such very rapid learning of new stimulus relations and its expression with little interference between successively learnt information, is a common theme of theories of hippocampal memory function (Bast, 2007;

Eichenbaum, 2004; O'Keefe & Nadel, 1979; O'Reilly & Rudy, 2001; Willshaw, Dayan, & Morris, 2015). Therefore, on theoretical grounds, DMP performance would be expected to be highly hippocampus-dependent. Indeed, as outlined in the Introduction, rodents' performance on the DMP watermaze task is highly sensitive to interference with hippocampal function, contrasting with performance on incremental watermaze tasks, which does not absolutely require hippocampal function. The remarkable similarity of DMP performance in humans and rodents revealed in the present study suggests similar underlying neuro-psychological mechanisms, including dependence on hippocampal function.

Interestingly, based on our correlational analysis, variability in latency and path length savings between trial 1 and 2 only partially predicts variability in search preference for the correct location, with variability in savings accounting for less than 20% (humans) or about 25% (rats) of the variability in search preference. This is in line with previous studies in rodents, which show that the neuro-psychological mechanisms underlying changes in these two measures can be partially dissociated. First, although latency and path length savings may also be affected by hippocampal manipulations, the search preference measure is more sensitive to hippocampal manipulations (Bast et al., 2009; McGarrity et al., 2017; Pezze & Bast, 2012). Second, in contrast to the latency and path length measures, search preference shows a slow gradual decay with increasing retention delay between trial 1 and 2, reaching chance levels within 24 h (da Silva et al., 2014). It has been suggested that such relatively rapid 'forgetting' is normally a key feature of the hippocampal memory system, balancing its capacity for very rapid learning and helping to minimize interference between memories (Hardt, Nader, & Nadel, 2013; Nonaka et al., 2017). Taken together, the partial correlation of latency and path length savings with search preference in the present study is consistent with findings from rodent studies suggesting that the search preference measure is more closely related to hippocampal function than savings. This probably reflects that a high search preference, with the subject

dwelling persistently in the correct location because they recognise the defining constellation of spatial cues, relies very strongly on a hippocampus-dependent rapidly formed allocentric representation of the goal location in relation to multiple distal spatial cues. In contrast, latency and path length savings may – at least partially – result from beacon strategies (i.e., heading towards a prominent landmark close to the goal), which do not require the hippocampus (Burešová, Krekule, Zahalka, & Bureš, 1985; Jacobs & Schenk, 2003; Morris, 1981; Schenk & Morris, 1985).

Our correlational analysis also revealed that, although both virtual DMP and mental rotation tests showed a similar sex difference (as discussed in 4.2), performance on both tests is only very weakly related, with search preference showing no significant correlation and latency and path length savings showing weak significant correlations ($r < 0.25$) to mental rotation scores in Replication 2, but no significant correlation in Replication 1. Previous studies comparing mental rotation performance to performance on incremental watermaze analogues reported mixed findings. Similar to our findings in Replication 2, Astur et al. (2004) reported a significant positive relation ($r < 0.41$) of participants' mental rotation scores to their performance on an incremental watermaze analogue, in terms of latency and path length measures, but not search preference, whereas Driscoll et al. (2005) found significant positive relations between mental rotation scores and performance on the watermaze analogue in terms of latency, path length and search preference measures ($r < 0.5$). Our findings suggest that mental rotation and 1-trial place learning performance on the virtual DMP task rely on largely distinct neuro-cognitive mechanisms. Consistently, as outlined in the Introduction and the preceding paragraph, rodent studies suggest that DMP performance depends on hippocampus-dependent memory mechanisms, whilst mental rotation has been associated with cortical visuo-motor regions, including parietal and motor cortical regions (Kosslyn, et al., 1998; Kosslyn et al., 2001).

The similarity of the new virtual DMP test to the rodent watermaze DMP test, and the high sensitivity of the latter to hippocampal manipulations (see Introduction and above), suggests that the new test offers the opportunity to probe hippocampal function with improved sensitivity as compared to standard incremental watermaze analogues. However, it is necessary to consider differences between watermaze experiments in rodents and virtual maze studies with human participants, which may have implications for the neurobiological mechanisms involved. First, our analyses of time spent in the zone surrounding the previous goal location during probe trials revealed that rats spent slightly more time searching at the previous location than would be expected by chance. In contrast, humans spent a negligible proportion of their time searching at the previous location, which is markedly less than would be expected by chance. Therefore, despite the similarity in performance noted earlier, rats, but not human participants, show pro-active interference based on memory of prior locations. This is likely to reflect that human participants 'understand' that the goal location moves every 4 trials, whereas rats appear to follow a 'win-stay' rule, searching in recently 'rewarded' goal locations; in line with the latter, during trial 1 to a new location on the DMP test, rats also show substantial search preference for the previous location, which is about twice as high as expected based on chance (McGarrity et al., 2017; Steele & Morris, 1999). Second, real world navigation in the watermaze, may rely on multi-modal information, including body-based self-motion information, which has been shown to contribute to place specific hippocampal neuron firing in rodents (Chen, King, Burgess, & O'Keefe, 2013). In contrast, joystick- or keyboard-controlled navigation in a virtual environment is mostly reliant on visual input (see Daugherty et al., 2016; Lavenex & Lavenex, 2010). In addition, relative to humans navigating in a virtual world, navigating in a watermaze is mildly aversive for rodents, which may have implications for the neurobiological mechanisms involved, although rodent studies support similar hippocampus-dependence of watermaze DMP protocols and food-reinforced dry-land DMP

tests (Bast, da Silva, & Morris, 2005; Nonaka et al., 2017). Despite these differences, previous studies using incremental place learning paradigms in virtual watermaze analogues indicate similar hippocampus-dependence to corresponding watermaze paradigms (Astur et al., 2002; Barkas et al., 2012; Goodrich-Hunsaker et al., 2010; Nedelska et al., 2012). Given the remarkably similar performance patterns of rodents and humans on the DMP paradigm, as revealed in the present study, we also expect similar hippocampus-dependence. Nevertheless, studies involving participants with partial hippocampal lesions (due to hippocampal sclerosis caused by epilepsy) are in preparation to verify the sensitivity of our paradigm to specific hippocampal damage.

4.2 Sex differences

Across both replications, male participants showed better 1-trial place learning performance on the virtual DMP test, as well as higher mental rotation scores. The higher mental rotation scores in male as compared to female participants replicate many previous findings (e.g., Astur et al., 2004; Peters et al., 1995; Voyer et al., 1995). The finding of robust sex differences in 1-trial place learning performance on the new virtual DMP test adds to previous findings that male, as compared to female, rodents or participants showed better incremental place learning in watermaze studies (Keeley et al., 2013; Perrot-Sinal, Kostenuik, Ossenkopp, & Kavaliers, 1996; Roof & Stein, 1999; Saucier et al., 2008), as well as in human studies using watermaze analogues (Astur et al., 1998; Astur et al., 2004; Driscoll et al., 2003; Woolley et al., 2010; Padilla et al., 2017). Importantly, the present study included two replications that both included 60 or more female and male participants and, therefore, were sufficiently powered (80%) to detect sex differences with an effect size corresponding to Cohen's d of about 0.5. Most previous studies comparing sex differences in incremental place

learning, especially in rodents, had substantially lower sample sizes, which increases the probability of both false negatives and false positives (Button et al., 2013). The limited statistical power probably explains why some studies failed to support significant sex differences in place learning, in terms of latency/path length measures or search preference, in the watermaze in rats (e.g., Bucci et al., 1995; Faraji et al., 2010) and in watermaze analogues in humans (Driscoll et al., 2005; Daugherty et al., 2016; Hamilton et al., 2009; Moffat & Resnick, 2002).

Given that rodent studies indicate that the DMP task is highly dependent on hippocampal function, better male performance on the virtual DMP task is consistent with evidence that male participants recruit the hippocampus to a greater extent than female participants during spatial navigation. Male participants navigating through a labyrinth-style virtual environment display (left) hippocampal activation, whilst female participants recruit right parietal and prefrontal cortex regions (Grön, Wunderlich, Spitzer, Tomczak, & Riepe, 2000). In addition, experiments conducted with rats (Rodríguez, Chamizo, & Mackintosh, 2011; Rodríguez, Torres, Mackintosh, & Chamizo, 2010) and humans (Sandstrom, Kaufman, & Huettel, 1998) suggest that male rats and male human participants preferentially navigate using boundary information rather than single landmark cues, whilst female rats and participants display the opposite preference. Given that the hippocampus has been implicated in processing boundary, but not single (proximal) landmark information (Doeller, King, & Burgess, 2008), these cue preferences are also consistent with the notion that male participants recruit hippocampal systems during navigation to a greater extent than female participants.

Whilst our study showed highly reliable sex differences in both the new virtual DMP task and the mental rotation test, with male participants performing on average better than female participants, the potential biological, psychological and social/environmental factors

underlying these sex differences (e.g., Baenninger and Newcombe, 1995; Levine et al., 2005; Hirnstein et al., 2014; Padilla et al., 2017; Picucci, Caffò, & Bosco, 2011) remain largely to be clarified. For example, while organizational and activational effects of the male sex hormone testosterone have been implicated in male advantages in mental rotation and place learning performance in virtual maze tasks (see reviews in Hirnstein et al., 2014; Nowak et al., 2014), it has also been demonstrated that gender stereotypes and whether testing took place in mixed- or same-sex groups can affect cognitive sex differences, including in mental rotation (Hirnstein et al., 2014), and that 'navigation' experience (as measured by how many of a list of local and national places participants had visited) affects sex differences in incremental place learning on a watermaze analogue (Padilla et al., 2017). Our study was not designed to address any of these factors; however, prompted by our participant demographics, which indicated that male participants were substantially more likely to regularly play computer games than female participants, we assessed if there was a relation between participants' report of regular computer gaming and virtual DMP and mental rotation performance. Previous studies have linked computer gaming to better performance on a number of cognitive tests, including the mental rotation test (Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Palaus, Marron, Viejo-Sobera, & Redolar-Ripoll, 2017; Terlecki & Newcombe, 2005), to an increase in hippocampal grey matter, and a shift from egocentric to allocentric navigation strategies (Kühn, Gleich, Lorenz, Lindenberger, & Gallinat, 2014). Consistent with these findings, our additional analyses suggest a positive relation between gaming experience and some aspects of mental rotation and virtual DMP performance, which may reflect that gaming improves performance or that the cognitive abilities that improve mental rotation and virtual DMP performance make it more likely to engage in regular computer gaming (see Boot et al., 2008). However, our analysis also suggested that gaming experience can only partially account for sex differences in mental rotation and virtual DMP performance, because mental rotation scores and search

preferences on the virtual DMP task were higher in male participants than in female participants in a subsample that did not report regular gaming experience.

5 CONCLUSIONS

Human participants on the new virtual DMP test show marked 1-trial place learning performance, which is remarkably similar to rodents' performance on the watermaze DMP test. Across two replications, male participants on average displayed better 1-trial place learning performance than female participants, adding to sex differences previously reported in both rodents and humans performing incremental place learning tasks on the watermaze or watermaze analogues, respectively. In addition, male participants on average performed better on the mental rotation test, consistent with previous studies. However, individual participants' performance on the virtual DMP test was largely unrelated to their mental rotation scores, suggesting that the virtual DMP test taps into neuro-cognitive mechanisms that are distinct from the mental rotation test. Given the remarkably similar performance pattern of humans and rodents on the DMP protocol, and the high sensitivity of DMP performance to hippocampal dysfunction in rodents, the new virtual DMP test may offer a highly sensitive tool to probe hippocampal function in human participants.

Acknowledgements

Data for the two replications reported in this paper were collected by Psychology students at the University of Nottingham for their final-year dissertation. We thank Hannah Frais, Tim

Rawlinson, Cherie Au-Yeung and Huachun Zhao for data collection in Replication 1, and Sebastian Leggett, Theo Gelernter, Emma Wilson, Syma Amin and Benson Lau for data collection in Replication 2. We also thank Michael Peters (University of Guelph) for providing the redrawn version of the Vandenberg and Kuse Mental Rotation Test.

References

Astur, R. S., Ortiz, M. L., & Sutherland, R. J. (1998). A characterization of performance by men and women in a virtual Morris water task:: A large and reliable sex difference. *Behavioural brain research, 93*(1), 185-190.

Astur, R. S., Purton, A. J., Zaniewski, M. J., Cimadevilla, J., & Markus, E. J. (2016). Human sex differences in solving a virtual navigation problem. *Behavioural brain research, 308*, 236-243.

Astur, R. S., Taylor, L. B., Mamelak, A. N., Philpott, L., & Sutherland, R. J. (2002). Humans with hippocampus damage display severe spatial memory impairments in a virtual Morris water task. *Behavioural brain research, 132*(1), 77-84.

Astur, R. S., Tropp, J., Sava, S., Constable, R. T., & Markus, E. J. (2004). Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. *Behavioural brain research, 151*(1), 103-115.

Ayaz, H., Allen, S., Platek, S., & Onaral, B. (2008). Maze Suite 1.0: A complete set of tools to prepare, present, and analyze navigational and spatial cognitive neuroscience experiments. *Behavior Research Methods, 40*(1), 353-359.

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex related differences in spatial and mathematical ability. *Learning and Individual Differences, 7*(4), 363-379.

Balcomb, F., Newcombe, N. S., & Ferrara, K. (2011). Finding where and saying where: developmental relationships between place learning and language in the first year. *Journal of Cognition and Development*, *12*(3), 315-331.

Bannerman, D., Good, M., Butcher, S., Ramsay, M., & Morris, R. (1995). Distinct components of spatial learning revealed by prior training and NMDA receptor blockade. *Nature, 378*(6553), 182.

Barkas, L., Redhead, E., Taylor, M., Shtaya, A., Hamilton, D. A., & Gray, W. P. (2012). Fluoxetine restores spatial learning but not accelerated forgetting in mesial temporal lobe epilepsy. *Brain*, *135*(8), 2358-2374.

Bast, T. (2007). Toward an integrative perspective on hippocampal function: from the rapid encoding of experience to adaptive behavior. *Reviews in the neurosciences, 18*(3-4), 253-282.

Bast, T., da Silva, B. M., & Morris, R. G. (2005). Distinct contributions of hippocampal NMDA and AMPA receptors to encoding and retrieval of one-trial place memory. *Journal of Neuroscience, 25*(25), 5845-5856.

Bast, T., Wilson, I. A., Witter, M. P., & Morris, R. G. (2009). From rapid place learning to behavioral performance: a key role for the intermediate hippocampus. *PLoS biology, 7*(4), e1000089.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta psychologica, 129*(3), 387-398.

Brady, A. M., & Floresco, S. B. (2015). Operant Procedures for Assessing Behavioral Flexibility in Rats. *Jove-Journal of Visualized Experiments*(96).

Brown, V. J., & Tait, D. S. (2016). Attentional Set-Shifting Across Species. *Current Topics in Behavioral Neurosciences, 28*, 363-395.

Bucci, D. J., Chiba, A. A., & Gallagher, M. (1995). Spatial learning in male and female Long-Evans rats. *Behavioral neuroscience, 109*(1), 180.

Buckley, M. G., Haselgrove, M., & Smith, A. D. (2015). The developmental trajectory of intramaze and extramaze landmark biases in spatial navigation: An unexpected journey. *Developmental psychology, 51*(6), 771.

Burešová, O., Krekule, I., Zahalka, A., & Bureš, J. (1985). On-demand platform improves accuracy of the Morris water maze procedure. *Journal of Neuroscience Methods, 15*(1), 63-72.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376.

Chapman, C. D., Benedict, C., & Schioth, H. B. (2017). Sex matters: Report experimenter gender. *Science, 356*, 916-917.

Chen, G., King, J. A., Burgess, N., & O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences, 110*(1), 378-383.

da Silva, B. M., Bast, T., & Morris, R. G. (2014). Spatial memory: behavioral determinants of persistence in the watermaze delayed matching-to-place task. *Learning & Memory, 21*(1), 28-36.

Daugherty, A. M., Bender, A. R., Yuan, P., & Raz, N. (2016). Changes in Search Path Complexity and Length During Learning of a Virtual Water Maze: Age Differences and Differential Associations with Hippocampal Subfield Volumes. *Cerebral Cortex, 26*(6), 2391-2401.

Daugherty, A. M., Yuan, P., Dahle, C. L., Bender, A. R., Yang, Y., & Raz, N. (2015). Path complexity in virtual water maze navigation: differential associations with age, sex, and regional brain volume. *Cerebral Cortex, 25*(9), 3122-3131.

de Hoz, L., Knox, J., & Morris, R. G. (2003). Longitudinal axis of the hippocampus: both septal and temporal poles of the hippocampus support water maze spatial learning depending on the training protocol. *Hippocampus, 13*(5), 587-603.

de Hoz, L., Moser, E. I., & Morris, R. G. (2005). Spatial learning with unilateral and bilateral hippocampal networks. *European Journal of Neuroscience, 22*(3), 745-754.

Doeller, C., King, J., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences of the United States of America, 105*(15), 5915-5920.

Driscoll, I., Hamilton, D. A., Petropoulos, H., Yeo, R. A., Brooks, W. M., Baumgartner, R. N., et al. (2003). The aging hippocampus: cognitive, biochemical and structural findings. *Cerebral cortex, 13*(12), 1344-1351.

Driscoll, I., Hamilton, D. A., Yeo, R. A., Brooks, W. M., & Sutherland, R. J. (2005). Virtual navigation in humans: the impact of age, sex, and hormones on place learning. *Hormones and Behavior, 47*(3), 326-335.

Fajnerová, I., Rodriguez, M., Levčík, D., Konrádová, L., Mikoláš, P., Brom, C., et al. (2014). A virtual reality task based on animal research–spatial learning and memory in patients after the first episode of schizophrenia. *Frontiers in Behavioral Neuroscience, 8*.

Faraji, J., Metz, G. A., & Sutherland, R. J. (2010). Characterization of spatial performance in male and female Long-Evans rats by means of the Morris water task and the ziggurat task. *Brain research bulletin, 81*(1), 164-172.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Folley, B. S., Astur, R., Jagannathan, K., Calhoun, V. D., & Pearlson, G. D. (2010). Anomalous neural circuit function in schizophrenia during a virtual Morris water task. *Neuroimage, 49*(4), 3373-3384.

Fritz, A. K., Amrein, I., & Wolfer, D. P. (2017). Similar reliability and equivalent performance of female and male mice in the open field and water-maze place navigation task. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 175*(3), 380-391.

Gazova, I., Vlcek, K., Laczó, J., Nedelska, Z., Hyncicova, E., Mokrisova, I., Sheardova, K. and Hort, J. (2012). Spatial navigation: A unique window into physiological and pathological aging. *Frontiers in Aging Neuroscience, 4*(16), 1-16.

Goodrich-Hunsaker, N. J., Livingstone, S. A., Skelton, R. W., & Hopkins, R. O. (2010). Spatial deficits in a virtual water maze in amnesic participants with hippocampal damage. *Hippocampus, 20*(4), 481-491.

Grön, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nature neuroscience, 3*(4), 404.

Hamilton, D. A., Akers, K. G., Johnson, T. E., Rice, J. P., Candelaria, F. T., & Redhead, E. S. (2009). Evidence for a shift from place navigation to directional responding in one variant of the Morris water task. *Journal of Experimental Psychology: Animal Behavior Processes, 35*(2), 271.

Hanlon, F. M., Weisend, M. P., Hamilton, D. A., Jones, A. P., Thoma, R. J., Huang, M., et al. (2006). Impairment on the hippocampal-dependent virtual Morris water task in schizophrenia. *Schizophrenia Research, 87*(1), 67-80.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences, 17*(3), 111-120.

Hirnstein, M., Coloma Andrews, L. & Hausmann, M. (2014) Gender-Stereotyping and Cognitive Sex Differences in Mixed- and Same-Sex Groups. *Archives of sexual Behavior, 43,* 1663-1673.

Hoh, T., Beiko, J., Boon, F., Weiss, S., & Cain, D. P. (1999). Complex behavioral strategy and reversal learning in the water maze without NMDA receptor-dependent long-term potentiation. *Journal of Neuroscience, 19*(10), RC2-RC2.

Hort, J., Laczo, J., Vyhnalek, M., Bojar, M., Bures, J., & Vlcek, K. (2007). Spatial navigation deficit in amnestic mild cognitive impairment. *Proceedings of the National Academy of Sciences, 104*(10), 4042-4047.

Hvoslef-Eide, M., Mar, A., Nilsson, S., Alsiö, J., Heath, C., Saksida, L. M., et al. (2015). The NEWMEDS rodent touchscreen test battery for cognition relevant to schizophrenia. *Psychopharmacology, 232*, 3853-3872.

Inglis, J., Martin, S. J., & Morris, R. G. (2013). Upstairs/downstairs revisited: spatial pretraining-induced rescue of normal spatial learning during selective blockade of

hippocampal N-methyl-d-aspartate receptors. *European Journal of Neuroscience, 37*(5), 718-727.

Jacobs, L. F., & Schenk, F. (2003). Unpacking the cognitive map: the parallel map theory of hippocampal function. *Psychological review, 110*(2), 285.

Jonasson, Z. (2005). Meta-analysis of sex differences in rodent models of learning and memory: a review of behavioral and biological data. *Neuroscience & Biobehavioral Reviews, 28*(8), 811-825.

Keeley, R. J., Tyndall, A. V., Scott, G. A., & Saucier, D. M. (2013). Sex difference in cue strategy in a modified version of the Morris water task: correlations between brain and behaviour. *PloS one, 8*(7), e69727.

Kosslyn, S. M., Digirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography. *Psychophysiology, 35*(2), 151-161.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews. Neuroscience, 2*(9), 635.

Kühn, S., Gleich, T., Lorenz, R. C., Lindenberger, U., & Gallinat, J. (2014). Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. *Molecular psychiatry, 19*(2), 265-271.

Lavenex, P. B., & Lavenex, P. (2010). Spatial relational learning and memory abilities do not differ between men and women in a real-world, open-field environment. *Behavioural brain research, 207*(1), 125-137.

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science, 16*(11), 841-845.

Leon, I., Cimadevilla, J. M., & Tascon, L. (2014). Developmental Gender Differences in Children in a Virtual Spatial Memory Task. *Neuropsychology, 28*(4), 485-495.

Logue, S. F., Paylor, R., & Wehner, J. M. (1997). Hippocampal lesions cause learning deficits in inbred mice in the Morris water maze and conditioned-fear task. *Behavioral neuroscience, 111*(1), 104.

McGarrity, S., Mason, R., Fone, K. C., Pezze, M., & Bast, T. (2017). Hippocampal neural disinhibition causes attentional and memory deficits. Cerebral Cortex. 27(9), 4447–4462.

Moffat, S. D., & Resnick, S. M. (2002). Effects of age on virtual environment place navigation and allocentric cognitive mapping. *Behavioral Neuroscience, 116*(5), 851.

Morris, R. G. M. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation, 12*(2), 239-260.

Morris, R. G. M. (2008). Morris water maze. *Scholarpedia, 3*(8), 6315.

Morris, R. G. M., Anderson, E., Lynch, G. A., & Baudry, M. (1986). Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. *Nature, 319*(6056), 774-776.

Morris, R. G. M., Hagan, J. J., & Rawlins, J. N. P. (1986). Allocentric spatial-learning by hippocampectomized rats - a further test of the spatial-mapping and working memory theories of hippocampal function. *Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology, 38*(4), 365-395.

Morris, R. G. M., Garrud, P., Rawlins, J., & O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature, 297*(5868), 681-683.

Morris, R. G. M., Schenk, F., Tweedie, F., & Jarrard, L. E. (1990). Ibotenate lesions of hippocampus and or subiculum - dissociating components of allocentric spatial-learning. *European Journal of Neuroscience, 2*(12), 1016-1028.

Morris, R. G. M., Steele, R. J., Bell, J. E., & Martin, S. J. (2013). N-methyl-d-aspartate receptors, learning and memory: chronic intraventricular infusion of the NMDA receptor antagonist d-AP5 interacts directly with the neural mechanisms of spatial learning. European Journal of Neuroscience, 37(5), 700-717.

Moser, M.-B., Moser, E. I., Forrest, E., Andersen, P., & Morris, R. (1995). Spatial learning with a minislab in the dorsal hippocampus. *Proceedings of the National Academy of Sciences, 92*(21), 9697-9701.

Nakazawa, K., McHugh, T. J., Wilson, M. A., & Tonegawa, S. (2004). NMDA receptors, place cells and hippocampal spatial memory. Nature reviews. *Neuroscience, 5*(5), 361.

Nakazawa, K., Sun, L. D., Quirk, M. C., Rondi-Reig, L., Wilson, M. A., & Tonegawa, S. (2003). Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. *Neuron, 38*(2), 305-315.

Nedelska, Z., Andel, R., Laczó, J., Vlcek, K., Horinek, D., Lisy, J., et al. (2012). Spatial navigation impairment is proportional to right hippocampal volume. *Proceedings of the National Academy of Sciences, 109*(7), 2590-2594.

Nonaka, M., Fitzpatrick, R., Lapira, J., Wheeler, D., Spooner, P. A., Corcoles-Parada, M., et al. (2017). Everyday memory: towards a translationally effective method of modeling the encoding, forgetting and enhancement of memory. *European Journal of Neuroscience, 46*(4), 1937-1953.

Nowak, N. T., Diamond, M. P., Land, S. J., & Moffat, S. D. (2014). Contributions of sex, testosterone, and androgen receptor CAG repeat number to virtual Morris water maze performance. *Psychoneuroendocrinology, 41*, 13-22.

O'Keefe, J., & Nadel, L. (1979). Précis of O'Keefe & Nadel's The hippocampus as a cognitive map. *Behavioral and Brain Sciences, 2*(4), 487-494.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychological review, 108*(2), 311.

Otnæss, M. K., Brun, V. H., Moser, M.-B., & Moser, E. I. (1999). Pretraining prevents spatial learning impairment after saturation of hippocampal long-term potentiation. *Journal of Neuroscience, 19*(24), RC49-RC49.

O'Carroll, C. M., Martin, S. J., Sandin, J., Frenguelli, B., & Morris, R. G. (2006). Dopaminergic modulation of the persistence of one-trial hippocampus-dependent memory. *Learning & Memory, 13*(6), 760-769.

Padilla, L. M., Creem-Regehr, S. H., Stefanucci, J. K., & Cashdan, E. A. (2017). Sex differences in virtual navigation influenced by scale and navigation experience. *Psychonomic Bulletin & Review, 24*(2), 582-590.

Palaus, M., Marron, E. M., Viejo-Sobera, R., & Redolar-Ripoll, D. (2017). Neural Basis of Video Gaming: A Systematic Review. *Frontiers in Human Neuroscience, 11*.

Panakhova, E., Buresova, O., & Bures, J. (1984). Persistence of spatial memory in the Morris water tank task. *International Journal of Psychophysiology, 2*(1), 5-10.

Perrot-Sinal, T. S., Kostenuik, M. A., Ossenkopp, K.-P., & Kavaliers, M. (1996). Sex differences in performance in the Morris water maze and the effects of initial nonstationary hidden platform training. *Behavioral Neuroscience, 110*(6), 1309-1320.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test-different versions and factors that affect performance. *Brain and Cognition, 28*(1), 39-58.

Pezze, M., & Bast, T. (2012). Dopaminergic modulation of hippocampus-dependent learning: blockade of hippocampal D1-class receptors during learning impairs 1-trial place memory at a 30-min retention delay. *Neuropharmacology, 63*(4), 710-718.

Picucci, L., Caffò, A. O., & Bosco, A. (2011). Besides navigation accuracy: Gender differences in strategy selection and level of spatial confidence. *Journal of Environmental Psychology, 31*(4), 430-438.

Robbins, T. W. (2002). The 5-choice serial reaction time task: behavioural pharmacology and functional neurochemistry. *Psychopharmacology, 163*(3-4), 362-380.

Rodriguez, P. F. (2010). Human navigation that requires calculating heading vectors recruits parietal cortex in a virtual and visually sparse water maze task in fMRI. *Behavioral neuroscience, 124*(4), 532.

Rodríguez, C. A., Chamizo, V., & Mackintosh, N. (2011). Overshadowing and blocking between landmark learning and shape learning: the importance of sex differences. *Learning & Behavior, 39*(4), 324-335.

Rodríguez, C. A., Torres, A., Mackintosh, N., & Chamizo, V. (2010). Sex differences in the strategies used by rats to solve a navigation task. *Journal of Experimental Psychology: Animal Behavior Processes, 36*(3), 395.

Roof, R. L. (1993). Neonatal exogenous testosterone modifies sex difference in radial arm and Morris water maze performance in prepubescent and adult rats. *Behavioural Brain Research, 53*(1), 1-10.

Roof, R. L., & Stein, D. G. (1999). Gender differences in Morris water maze performance depend on task parameters. *Physiology & Behavior, 68*(1), 81-86.

Roof, R. L., Zhang, Q., Glasier, M. M., & Stein, D. G. (1993). Gender-specific impairment on Morris water maze task after entorhinal cortex lesion. *Behavioural Brain Research, 57*(1), 47-51.

Sandstrom, N. J., Kaufman, J., & Huettel, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Cognitive Brain Research, 6*(4), 351-360.

Saucier, D., & Cain, D. P. (1995). Spatial learning without NMDA receptor-dependent long-term potentiation. *Nature, 378*(6553), 186-189.

Saucier, D., Shultz, S., Keller, A., Cook, C., & Binsted, G. (2008). Sex differences in object location memory and spatial navigation in Long-Evans rats. *Animal Cognition, 11*(1), 129-137.

Schenk, F., & Morris, R. (1985). Dissociation between components of spatial memory in rats after recovery from the effects of retrohippocampal lesions. *Experimental Brain Research, 58*(1), 11-28.

Steele, R., & Morris, R. (1999). Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP5. *Hippocampus, 9*(2), 118-136.

Sutherland, R. J., Whishaw, I. Q., & Kolb, B. (1983). A behavioural analysis of spatial localization following electrolytic, kainate-or colchicine-induced damage to the hippocampal formation in the rat. *Behavioural Brain Research, 7*(2), 133-153.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles, 53*(5), 433-441.

Tsien, J. Z., Huerta, P. T., & Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor–dependent synaptic plasticity in spatial memory. *Cell, 87*(7), 1327-1338.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599-604.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Whishaw, I. Q. (1985). Cholinergic receptor blockade in the rat impairs locale but not taxon strategies for place navigation in a swimming pool. *Behavioral Neuroscience, 99*(5), 979-1005.

Whishaw, I. Q., & Jarrard, L. E. (1996). Evidence for extrahippocampal involvement in place learning and hippocampal involvement in path integration. *Hippocampus, 6*(5), 513-524.

Williams, C. L., & Meck, W. H. (1991). The organizational effects of gonadal steroids on sexually dimorphic spatial ability. *Psychoneuroendocrinology, 16*(1), 155-176.

Willshaw, D., Dayan, P., & Morris, R. (2015). Memory, modelling and Marr: a commentary on Marr (1971)'Simple memory: a theory of archicortex'. *Philosophical Transactions of the Royal* Society *B, 370*(1666), 20140383.

Woolley, D. G., Vermaercke, B., de Beeck, H. O., Wagemans, J., Gantois, I., D'Hooge, R., et al. (2010). Sex differences in human virtual water maze performance: Novel measures reveal the relative contribution of directional responding and spatial knowledge. *Behavioural Brain Research, 208*(2), 408-414.

Figure legends

Figure 1: The new virtual delayed-matching-to-place (DMP) test. A) Schematic diagram of the virtual environment used. Black squares indicate 8 possible goal locations, and dotted squares represent the search zones used for the analysis of search preference for the 'correct zone' containing the correct location during probe trials. Letters N, E, S, W represent the four start locations at which trials began, and symbols represent the objects that were located beyond the walls of the circular arena to serve as distal spatial cues. B) A schematic representation of the DMP paradigm, with filled circles indicating the hidden goal location within the circular environment. Participants were instructed to find a hidden goal (William the Worm). The goal remained in the same location for only four consecutive trials and was then moved to a new location for another block of four trials. Therefore, the paradigm requires the repeated, rapid learning of new goal locations. Altogether, participants completed six blocks of four trials with six different locations. C) Views of the environment from a participant's perspective.

Figure 2: Illustrative examples of the paths taken by participants across trials 1-4 to a given goal location (black square). Participants typically searched the environment systematically on trial 1, adopting either a circular search strategy (top panel), a zig-zag strategy (middle panel), or a hybrid of both strategies (bottom panel). Following trial 1, participants typically moved more directly towards the goal location, reflecting 1-trial place learning.

Figure 3: Key performance measures (±SEM) on the new virtual DMP test in male and female participants. Means (± SEM) of latencies (top panel) and path lengths (middle panel) across the 4 trials, and of the percentage time in the correct zone during probe trials (bottom panel) are shown for Replication 1 (left) and Replication 2 (right). The stippled line in the bar graphs (bottom panel) indicates chance level (12.5%). In both replications, male and female participants displayed significant latency and path length savings between trial 1 and 2,

reflecting 1-trial place learning; however, the savings in male participants were significantly greater than the savings in female participants. Moreover, in both replications, male and female participants spent more time in the correct zone during probe trials than would be expected by chance, supporting significant 1-trial place learning; however, male participants spent a significant greater proportion of time in the correct zone relative to female participants on both probe trials.

Figure 4: Correlation between search preference and latency savings on the virtual DMP test. Mean percentage time in the correct zone averaged across both probe trials plotted against individual latency savings made between trial 1 and 2 for Replication 1 (left panel) and 2 (right panel). In both replications, there was a significant weak to moderate correlation between search preference and latency savings.

Figure 5: Performance of adult male Lister hooded rats on the watermaze DMP test (re-analysis of data from Bast et al., 2009). Mean latencies (±SEM) for rats to find the submerged platform (top panel), mean percentage time in the correct zone (+SEM) during probe trial (middle panel; stippled line indicates chance level), and mean percentage time in the correct zone plotted against individual latency savings (bottom panel) are shown for n = 100 rats. The rat data are remarkably similar to the human data, showing marked latency savings between trial 1 and 2, a marked search preference for the correct zone during probe trials, as well as a significant moderate correlation between latency savings and search preference.

Figure 6: Mental rotation in male and female participants. Mean (+SEM) scores on the mental rotation test for male and female participants in Replication 1 (left panel) and 2 (right panel). In both replications, male participants displayed significantly greater mental rotation scores than female participants.

Figure 7: Correlations between mental rotation scores and key performance measures on the virtual DMP test. Correlations between mental rotation scores and time spent in the correct zone (top panel), latency savings (middle panel), and path length savings (bottom panel) are shown for Replication 1 (left) and Replication 2 (right). There were no consistent correlations between these measures. In Replication 1, mental rotation scores did not correlate with search preference, latency savings, or path length savings. In Replication 2, there was again no correlation between mental rotation scores and search preference, although both latency and path length savings were weakly, but significantly, correlated with mental rotation scores.